PAPERS PRESENTED AT A ONE DAY WORKSHOP ON STATISTICAL APPLICATIONS IN MINING

Held at Murdoch University

October 1st 1996

EDITORS:

B. R. CLARKE

and

I. W. WRIGHT

Printed by Murdoch University Press

Western Australia

ISBN: 0-86905-524-0

Testing for Outliers in Gold Assay Data; Unweighted and Weighted Samples

Dr. Brenton R. Clarke

Department of Mathematics and Statistics,
Murdoch University, Murdoch,
Western Australia 6150

Facsimile (+61 9) 360 6332
Telephone (+61 9) 360 2578

TESTING FOR OUTLIERS IN GOLD ASSAY DATA; UNWEIGHTED AND WEIGHTED SAMPLES

by
Dr Brenton R Clarke
Department of Mathematics and Statistics
Murdoch University

My involvement in this workshop has had its origins in a piece of consulting work with a gold mining company, Company A, in 1989. I received a fax from a concerned metallurgist who was responsible for overseeing the contracts and processing for ore crushed at the Company A Mine. The metallurgist was assessing the mean recoverable gold content from ore that was being purchased from Company B. Payment was on the basis of the assessed mean recoverable gold content less 2 g/t treatment charge. The problem is to determine the "best estimate" of the mean recoverable gold content.

On the occasion in question a tonnage weighted arithmetic mean was specified as the basis for grade determination ($12.94g/t$). It had been suggested to the metallurgist following the contract, that, as gold distributions are frequently skewed, the geometric mean will give a more accurate estimate of the true value. He was aware that the geometric mean will reduce the significance of very high values, but he wondered whether it tends to underestimate the true mean value? He doubted whether the Joint Venture partners would agree to using the geometric mean if that was the case.

For the data in question (see Appendix) which were in ordered value the following statistics were evaluated

Tonnage Weighted Arithmetic Mean	$= \mathbf{12.94}\ g/t$
Arithmetic Mean	$= \mathbf{13.2396}\ g/t$
Geometric Mean	$= \mathbf{11.8809}\ g/t$

The metallurgist noted, as can we, that the difference between the arithmetic mean and the geometric mean was a significant component of the treatment charge. If the geometric mean value was the more "correct" value, Company A treated the Joint Venture ore for nothing!

When are the above estimates optimal and what are they estimating? What other estimates could we use?

Firstly, consider the historical perspective in advocating the arithmetic mean. Gauss (1777-1855) argued for the normal distribution based on the popular use of the mean (arithmetic). The name of Gauss is often associated with this bell shaped curve which takes the formula

$$f(x;\mu,\sigma^2) = \frac{1}{\sqrt{2\pi}\sigma}\, e^{-(x-\mu)^2/2\sigma^2} \quad \begin{array}{l} -\infty < x < \infty; \\ -\infty < \mu < \infty \ ; \ \sigma > 0 \end{array}$$

The mean of an observation from this density curve is μ, and the standard derivation is σ. Let us assume our sample of n independent observations, denoted by $X_1, ..., X_n$, come from this distribution so that their joint distribution is the product of the density curves evaluated at the sample points

$$L_n(\mu,\sigma^2) = \prod_{i=1}^{n} f(X_i;\mu,\sigma^2) \tag{1}$$

Gauss pointed out that if the data had a normal distribution, then $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ is the optimal estimator. This we argue for today based on the fact that if we examine (1) which is known as the "likelihood" of the observations $X_1, ..., X_n$ given μ, σ^2 then the maximising value for the likelihood is obtained when

$$\frac{\partial^2 \ell n\ L}{\partial\mu\partial\sigma^2} = 0$$

which gives solutions, estimates for μ and σ^2 respectively

$$\overset{\vee}{\mu}{=}\bar{X} \quad ; \quad \overset{\vee}{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$$

But even a contemporary of Gauss wrote:

"*For example, there are certain provinces of France where to determine the mean yield of a property of land, there is a custom to observe this yield during twenty consecutive years, to remove the strongest and weakest yield and then to take one eighteenth of the sum of the others.*" ... *Huber (1972)*

The practice of throwing large and small observations away is to help robustify the estimate of the mean of a population. The French farmers

were using what is known as a 5% trimmed mean. Taking this approach for the sample of 48 observations ordered from smallest to largest taken from the gold assays of the Joint Venture ore would leave, for an approximate 5% trimmed mean, 44 observations after discarding the two largest and two smallest observations. This leads to an estimate

$$T_2 = \mathbf{12.24} \tag{2}$$

Trimming is actually 4.16% from each tail or 8.33% overall. Note the difference with the arithmetic mean!

To extend this argument further, a natural question to ask is how many observations should we trim? If the sample is long-tailed, thus appearing to come from a long-tailed distribution, one would choose to trim $g > 0$ values from each tail if indeed the data were from a symmetric distribution. In the limiting case of discarding almost all the samples, to quote Tukey and McLaughlin (1963)

"*the cost will not be paid in terms of the stability of* T_g*, which cannot be made worse than the stability of the median - which is excellent in sampling from long-tailed distributions - but rather in terms of the stability of the estimate of the variability of* T_g"

Tukey and McLaughlin observed that the variance $\sigma^2(T_g)$ is minimized for $g = 0$ when $T_0 = \overline{X}$. They proposed an estimator of the mean where one chooses $\widehat{g} \leq G(n)$ so that an estimate of the variability of T_g is minimized. That is, for a suitable estimate of variability $\widehat{\sigma}^2(T_g)$, the number of observations chosen to trim from each tail, $\widehat{g}$, satisfies

$$\widehat{\sigma}^2(T_{\widehat{g}}) = \min_{0 \leq g \leq G(n)} \widehat{\sigma}^2(T_g) \tag{3}$$

For the resulting $\widehat{g}$ the estimator of the mean μ of the population is given by $T_{\widehat{g}}$. A typical choice of $G(n) = [n/4]$, see Clarke (1994).

This idea is linked to that of recognising and casting out spurious observations. However, the idea of trim,trimming both tails equally may mean that some observations are thrown out unnecessarily. Moreover, it has been shown (cf. Clarke (1994)) that even if the data are normal, for the accepted

choice of $\widehat{\sigma}^2(T_g)$, the choice of $\widehat{g}$ can more **frequently** range over the region

$$1 \leq \widehat{g} \leq G(n), \tag{4}$$

as opposed to the optimal preferred $g = 0$. This suggests that if the data are normal, the adaptive trimmed mean will unnecessarily trim observations, and we may be far from our optimal estimator (assuming normality) of $T_0 = \overline{X}$.

To avoid complications of trimming unnecessarily and to only trim the unusual observations that may be only on one side of the data, as may be the case in asymmetric departures from normality, it is proposed in Clarke (1994) to use an adaptive version of the trimmed likelihood estimator of Bednarski and Clarke (1993). First let us introduce some notation. Let S_n be the set of all possible choices of $(n-g)$ observations from the n observations $\{X_1, ..., X_n\}$. Denote the particular set of observations $\{X_{j_1}, ..., X_{j_{n-g}}\}$ and the parameter $\widetilde{\mu}_g$ to be the observations and parameters which maximise the trimmed likelihood

$$L_{n-g}(X_{j_1}, ..., X_{j_{n-g}}, \widetilde{\mu}_g, \sigma^2) = \max_{\{X_{i_1}, ..., X_{i_{n-g}}\} \in Sn} \max_{\mu} L_{n-g}(X_{i_1}, ..., X_{i_{n-g}}; \mu, \sigma^2) \tag{5}$$

Following Tukey and McLaughlin's example, the estimator of Clarke (1994) estimates $\widetilde{g}$ so that

$$\sigma_*^2(\widetilde{g}) = \min_{0 \leq g \leq [n/2]} \sigma_*^2(g) \tag{6}$$

where $\sigma_*^2(g)$ is a suitable estimate of variance of $\widetilde{\mu}_g$.

The resulting estimator is $\widetilde{\mu}_{\widetilde{g}}$ will be

$$\widetilde{\mu}_{\widetilde{g}} = \frac{1}{n - \widetilde{g}} \sum_{\ell=1}^{n-\widetilde{g}} X_{j_\ell} \ , \tag{7}$$

which is the mean of the observations excluding those trimmed in the likelihood of

$$\{X_1, ..., X_n\} \setminus \{X_{j_1}, ..., X_{j_{n-\widetilde{g}}}\} \tag{8}$$

The observations in (8) are regarded as outliers.

Remark: We have abbreviated notation in that $\{X_{j_1}, ..., X_{j_{n-g_1}}\}$ may not be the same as $\{X_{j_1}, ..., X_{j_{n-g_2}}\}$ when calculated in equation (5).

The gold data are analysed in Clarke (1994) using this method, which incidentally gives $\widetilde{g} = 1$ outlier. The estimator

$$\widetilde{\mu}_{\widetilde{g}} = \mathbf{12.21} \tag{9}$$

corresponds to the mean of the observations with the observations 61.5 trimmed from the sample.

If we are seeking graphical reasons why we might trim this largest observation then consider the histogram of the data. This clearly shows the observation that is outlying see figure 1. A normal probability plot also would show the outlying observation. See figures 2,3. If we are going to go with the assumption of normality then it is only reasonable that the observation 61.5 should be trimmed. see figures 1,2,3

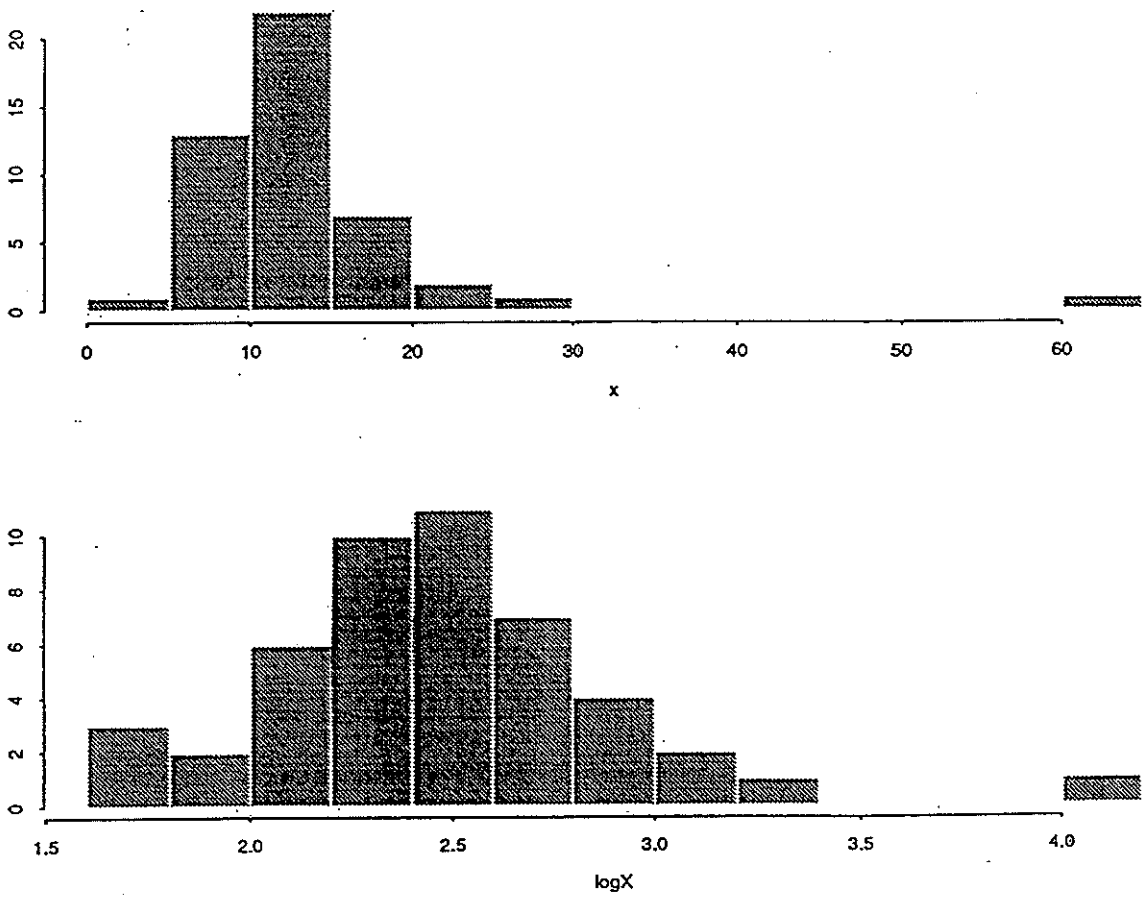

Figure 1: Histograms of Raw and Logged Gold Assays

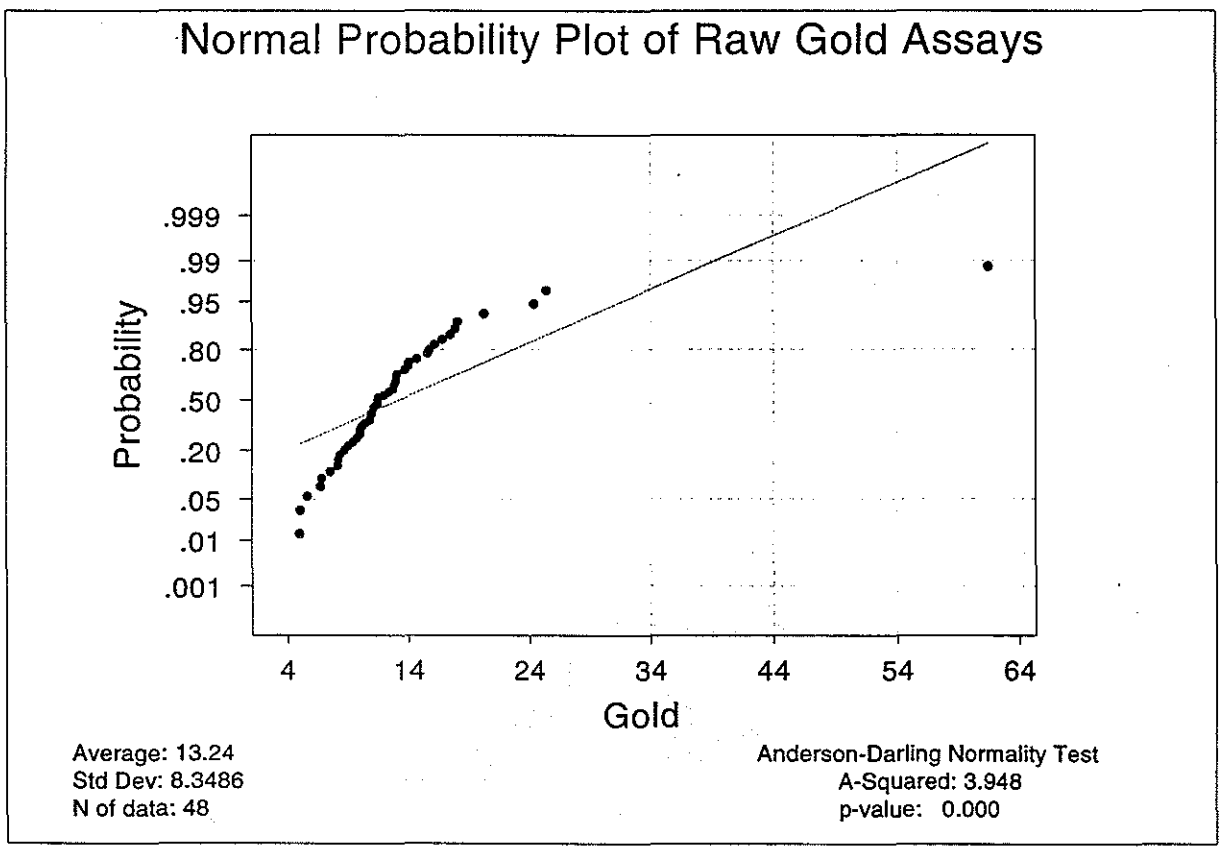
Normal Probability Plot of Raw Gold Assays
Probability
.999
.99
.95
.80
.50
.20
.05
.01
.001
4
14
24
34
44
54
64
Gold
Average: 13.24
Std Dev: 8.3486
N of data: 48
Anderson-Darling Normality Test
A-Squared: 3.948
p-value: 0.000

Figure 2:

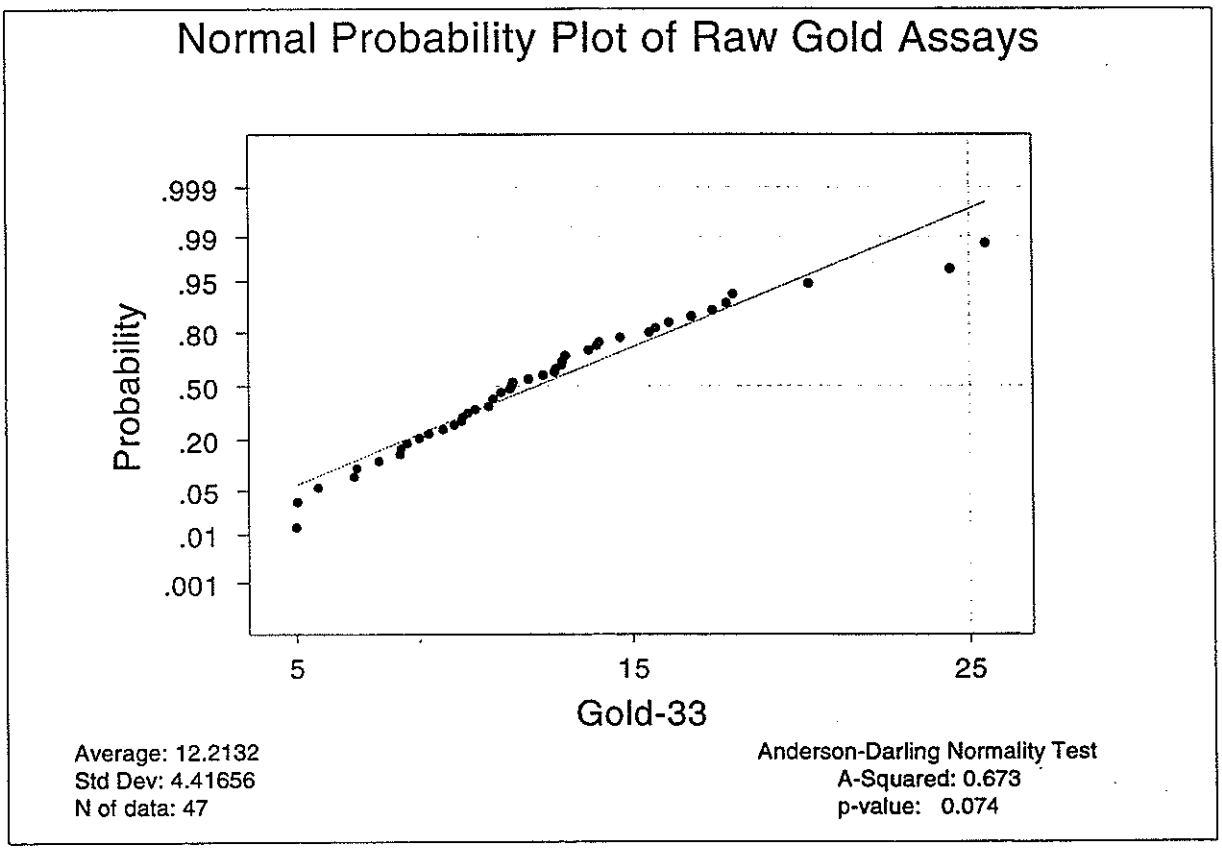

Figure 3: This is the normal probability plot of the raw data excluding the 33rd observation from the original sample which corresponds to removing the largest observation.

However, there is a suggestion that the data are skewed. Indeed Krige (1951, 1962) describes gold values as frequently being described by a log normal distribution. The corresponding density function is then

$$f(y;\mu,\sigma^2) = \frac{1}{\sqrt{2\pi}\sigma\ y}\ e^{-\frac{1}{2}[(\ell n\ y-\mu)/\sigma]^2} \qquad y>0 \tag{10}$$

Here if Y has the log normal density, then $X = \ell n(Y)$ is a normal variable with mean μ and variance σ^2. The estimator for μ is $\overline{X}$, i.e., the mean of the logged values, and re-exponentiating gives the geometric mean

$$GM = Y_1^{1/n}...Y_n^{1/n} \tag{11}$$

assuming $Y_1, ..., Y_n$ are the original observations.

We know $\overline{X}$ is the optimal estimator for μ if the data $\{Y_i\}$ are lognormal, and hence GM is the optimal estimator for e^{μ}. But e^{μ} is simply the median of the lognormal distribution which is a skew distribution and it is therefore argued that we should, since we are interested in the return, use the mean of the lognormal distribution which in fact can be shown to be

$$e^{\mu+\frac{\sigma^2}{2}} \tag{12}$$

Sichel (1952,1966) proposed the estimator for the mean

$$[GM] \quad \times \quad \left[\begin{array}{c} \textit{factor dependent on sample} \\ \textit{variance of logged values} \\ \textit{and sample size } n \end{array}\right] \tag{13}$$

Researchers can fossick for the details. Krige (1978) Table 1 gives the factor which for this data ≈ 1.1 so that the estimator for the mean

$$GM \times factor = 11.88 \times 1.1 = \mathbf{13.07} \tag{14}$$

This is much closer to the arithmetic mean of the data. Please note, we do not use the geometric mean by itself. Nevertheless, it is clear also that the assumption of logged values following a normal distribution should be checked. Clearly, from the normal probability plot of logged values in figure 4 the largest observation is an outlier. This can also be observed from the

histogram, figure 1.
The adaptive trimmed likelihood mean estimator when applied to the logged data also identifies that one observation is outlying. The resulting estimate after culling the one observation gives the geometric mean

$$GM = 11.47 \tag{15}$$

The resulting variance estimate, again after culling the largest observation, leads to a factor from Table 1 of Krige (1978) of 1.07 so that the estimate for the mean is

$$GM \times [factor] = 11.47 \times 1.07 = \mathbf{12.27} \tag{16}$$

This is in fair agreement with the arithmetic mean with the largest observation removed from the data, that is, with the estimate given by $\widetilde{\mu}_{\widetilde{g}}$ above.

Finally, let us examine the estimate based on a weighted mean. When is this optimal and what is the reason for using it? The gold assays were taken on different tonnages of ore. It is accepted in the literature on gold mining, and makes intuitive common sense, that the variability of the gold assay values is inversely proportional to the weight of ore from which they are derived. Here

$$var[X_i] \quad \alpha \quad 1/w_i \qquad\qquad i = 1, ..., 48 \tag{17}$$

where w_i is the dry weight. Presumably, the more ore from which you take your assay from, the more accurate is the estimate . That is the density of the i'th observation is assumed to be

$$f(X_i; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi}} \; \frac{1}{\sigma/\sqrt{w_i}} \; e^{-\frac{1}{2}(X_i-\mu)^2/(\sigma^2/w_i)} \tag{18}$$

Here σ^2 is the constant of proportionality in (17). The observations are assumed independent, whereupon the classical maximum likelihood estimator for the mean is

$$\widehat{\mu} = \sum_{i=1}^{n} \frac{w_i X_i}{\left(\sum_{j=1}^{n} w_j\right)} \tag{19}$$

This is the tonnage weighted arithmetic mean on which the contract is based. The estimate of variance is

$$\widehat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} w_i (X_i - \widehat{\mu})^2 \tag{20}$$

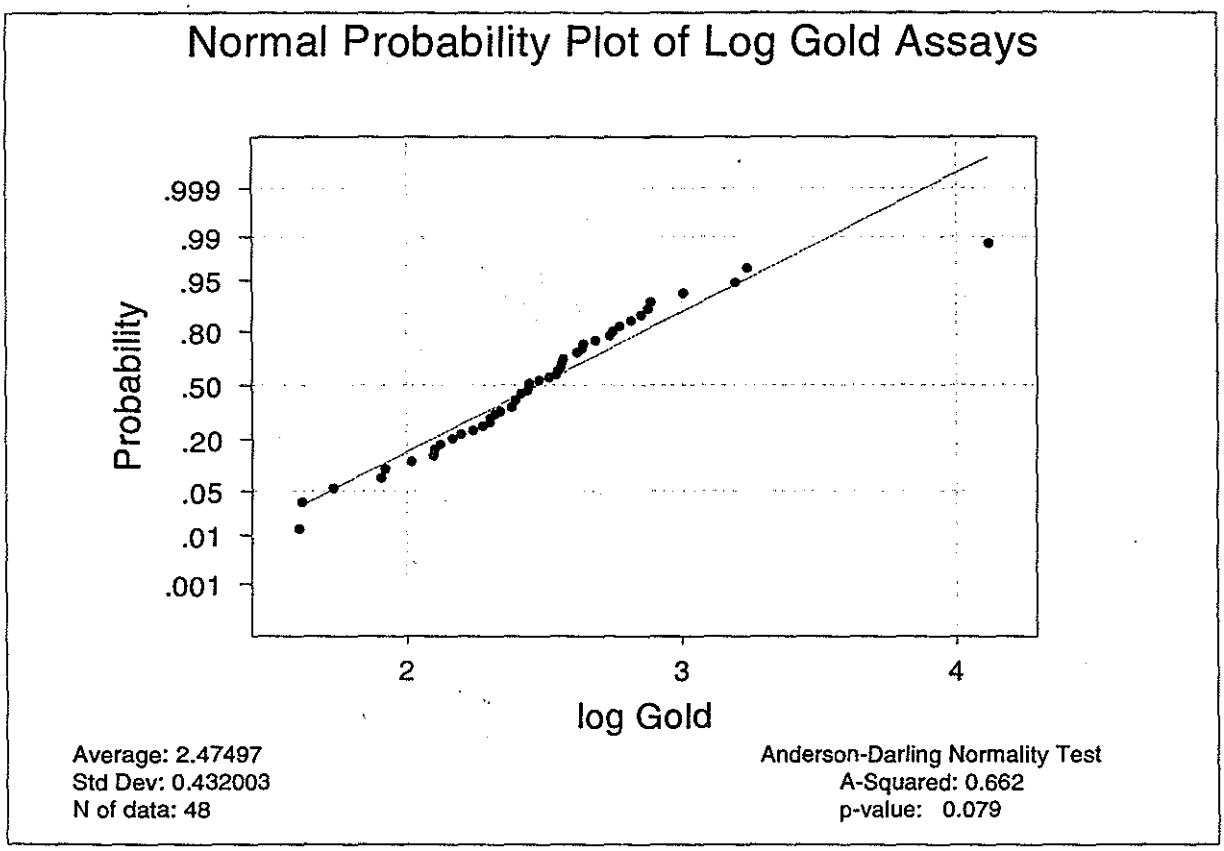
Normal Probability Plot of Log Gold Assays
Probability
.999
.99
.95
.80
.50
.20
.05
.01
.001
2
3
4
log Gold
Average: 2.47497
Std Dev: 0.432003
N of data: 48
Anderson-Darling Normality Test
A-Squared: 0.662
p-value: 0.079

Figure 4:

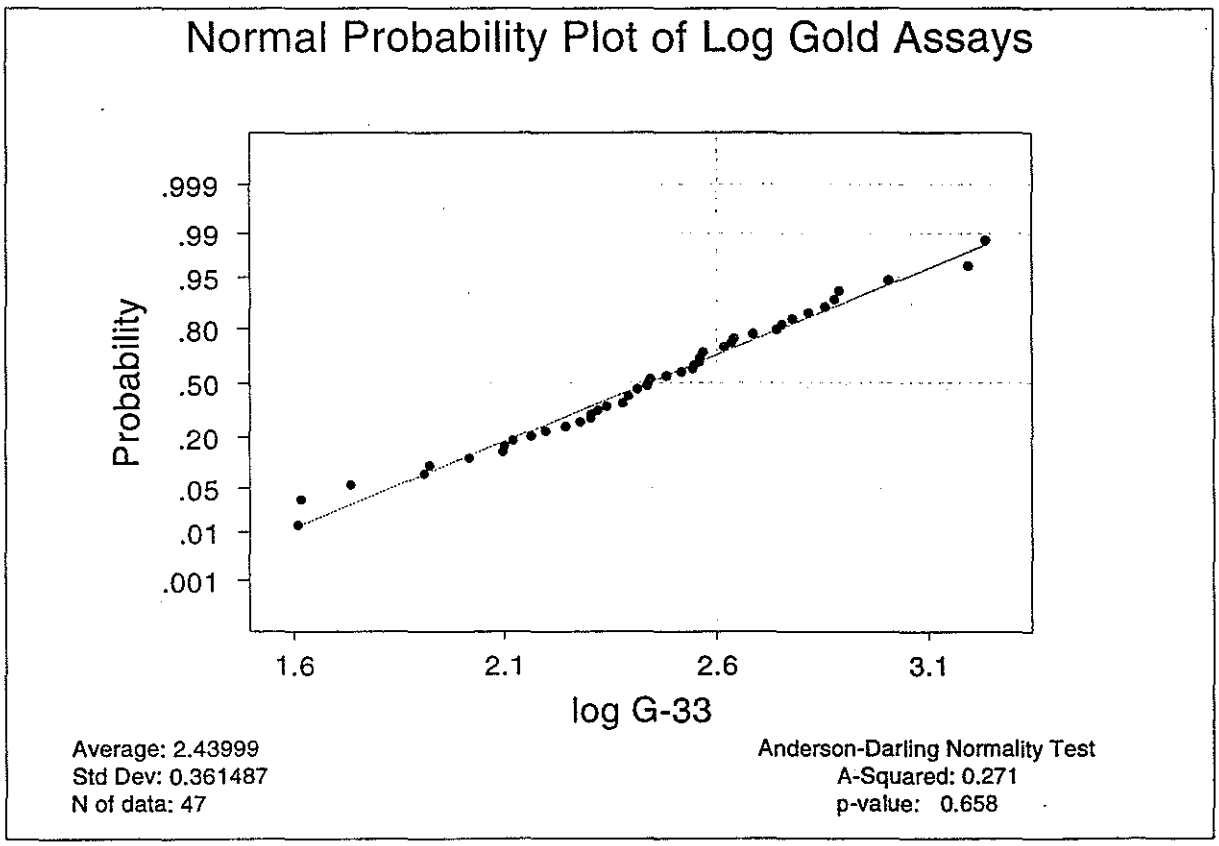

Figure 5: This is the normal probability plot of the logged data excluding the 33rd observation from the original sample, which corresponds to removing the largest observation.

If the assumed normal distribution is correct the location and variance estimators are jointly sufficient and as the location estimator is also unbiased it is a minimum variance unbiased estimator.

How can one formulate robust estimates of parameters μ and σ^2 in the case of weights? This question has been partly addressed in Armstrong (1994). There is for example what is known as Huber's Proposal 2 system of estimating equations where taking $\psi(X) = \max(-k, \min(X, k))$ and choosing $a(k)$ so that

$$\int \psi^2(z) \frac{1}{\sqrt{2\pi}} e^{-z^2/2} \, dz = a(k)$$

the estimating equations for location and scale assuming weights become

$$\sum_{i=1}^{n} \psi\left(\frac{(X_i-\mu)\sqrt{w_i}}{\sigma}\right)\sqrt{w_i} = 0$$
$$\sum_{i=1}^{n} \psi^2\left(\frac{(X_i-\mu)\sqrt{w_i}}{\sigma}\right) = n\, a(k)$$

These equations have as their basis that they mimic the maximising likelihood equations, but bound the influence of outlying values. Choosing $k = 1.5$ which corresponds to a 95% efficient estimator at the normal model, the resulting estimate gives

$$\widehat{\mu}_{\text{Proposal 2}} = \mathbf{12.15}$$

Armstrong also attempted to redefine the adaptive trimmed likelihood estimator for the case of unequal weights. However, given for a fixed trimming number g , while the trimmed likelihood estimator of the mean for weighted data,

$$\widehat{\mu}_w(g) \tag{21}$$

can be defined, there is no obvious candidate for $\widehat{\sigma}^2_{*w}(g)$ which would be the estimated asymptotic variance of $\widehat{\mu}_w(g)$. Armstrong substitutes the asymptotic variance formula similar to the case of equal weights although this gave two minimising values for $\widehat{\sigma}^2_{*w}(g)$, one at $g = 1$, the other at $g = 3$. Choosing $g = 1$ gives

$$\widehat{\mu}_w(1) = \mathbf{12.41} \tag{22}$$

where the observation trimmed corresponds to the largest observation, and $g = 3$ gave an estimate

$$\widehat{\mu}(3) = 11.83$$

The diagnostic approach of plotting normal probability plots is not readily available for not identically distributed data, that is, the case of data with unequal weights. However, based on the study of the data in the case of equal weights, we may well ask, is there a single outlier in the sample, can we identify it and test for it. This is the subject of joint collaborative work with Professor Lewis of the Centre for Statistics, University of East Anglia, and tentative results show that using accepted methodology for outlier analysis a new test can be defined which shows the observation corresponding to the largest gold assay value is outlying. Then the maximum likelihood estimator based on the remaining data yield

$$\widehat{\mu} = \mathbf{12.41}, \tag{23}$$

this coinciding with the adaptive trimmed likelihood estimator, $\widehat{\mu}_w(1)$

Postscript: Since the preparation and delivery of this manuscript the work with Toby Lewis suggests that with unequal weights and assuming logged values follow a normal probability model, there may be more than one outlier. However, consideration of models other than normal may deem there to be no outliers. This work will be published elsewhere.

Armstrong, N.J.(1994), *Robust Analysis and Identification of Outliers for Data Incorporating Known Weights*, Honours thesis in Mathematics, Murdoch University, Murdoch, W.A., Australia.

Bednarski, T. and Clarke, B.R.(1993), *Trimmed likelihood estimation of location and scale of the normal distribution*, Australian Journal of Statistics, **35**, 141-153.

Clarke, B.R.(1994), *Empirical evidence for adaptive confidence intervals and identification of outliers using methods of trimming*, Australian Journal of Statistics, **36**, 45-58.

Huber, P.J.(1972), *Robust statistics : a review*, Ann. Math. Statist. **43**, 1041-1067.

Krige, D.G.(1951), *A statistical approach to some mine valuation and allied problems on the Witwatersrand*, M.Sc.(Eng) thesis, University of the Witwatersrand, Johannesburg.

Krige, D.G.(1962), *Statistical applications in mine valuation*, J. Inst. Mine Survey. S.Afr., **12(2)**, 45-84, **12(3)**, 95-136.

Krige, D.G.(1978), *Lognormal-de Wijsian Geostatistics for Ore Evaluation*, South African Institute of Mining and Metallurgy, Johannesburg.

Sichel, H.S.(1952), *New methods in the statistical evaluation of mine sampling data*, Bull. Inst. Min. Metall., Lond., June 1952, 261-288.

Sichel, H.S.(1966), *The estimation of means and associated confidence limits for small samples from lognormal populations.* In Proc. Symp. on Mathematical Statistics and Computer Applications in Ore Valuation, 106-123, S. Afr. Inst. Min. Metall., Johannesburg.

Tukey, J.W. and McLaughlin, D.H.(1963), *Less vulnerable confidence and significance procedures for location based on a single sample : Trimming/Winsorization 1*, Sankhya Ser. A, **25**, 331-352.

APPENDIX: *Raw Gold Assays in grammes per ton were taken from the sample of 48 observations provided by Company A*

5.00,	5.04,	5.67,	6.75,	6.84,	7.52,	8.15,	8.18,	8.35,	8.72,
9.00,	9.43,	9.76,	10.00,	10.02,	10.18,	10.39,	10.80,	10.94,	10.94,
10.94,	11.17,	11.43,	11.46,	11.52,	11.97,	12.40,	12.74,	12.78,	12.94,
12.96,	13.05,	13.73,	13.96,	14.04,	14.67,	15.53,	15.71,	16.10,	16.76,
17.40,	17.81,	18.00,	20.25,	24.45,	25.50,	61.5.			